April 2018

Rates and Patterns of Evolution in a Duplicated Genome in the Family Catostomidae

Megann Schmidt
University of Northern Iowa, schmimax@uni.edu

Rates and Patterns of Evolution in a Duplicated Genome in the Family Catostomidae

Megann Schmidt, Zachary Sperstad, and Peter Berendzen

Department of Biology

INTRODUCTION:

Whole genome duplication (WGD) is a process by which the entire genome of an organism is copied, leading to polyploidy. WGD events result in duplicate copies of genes that are subject to a variety of evolutionary forces. If there is purifying selection on the gene, all copies will retain their initial function, but under other modes of selection, the duplicated genes can develop a new function or divide the ancestral function among the duplicates. In the absence of selection, one copy can lose functionality and become a pseudogene (Fig.1). It has been widely hypothesized that WGD events are a major driving force behind diversification. However, recent studies have not supported this hypothesis in animals and have commonly observed retention of the initial gene state or loss of duplicate gene function through degenerate mutation.

Genome duplication occurs across a wide range of animals, but is particularly common in fish and amphibians. This is likely due to external fertilization in freshwater environments, exposing the zygotes to temperature stress, along with other factors such as communal breeding and large numbers of gametes.

The family Catostomidae, the suckers, is hypothesized to have evolved from a single tetraploid ancestor resulting from a WGD event. Based on early studies using isozyme data, it has been hypothesized that within the group the subfamily Ictiobinae, retained initial gene states more frequently than its sister group, the subfamily Catostomimae. The objective of this study is to test this hypothesis using genomic data.

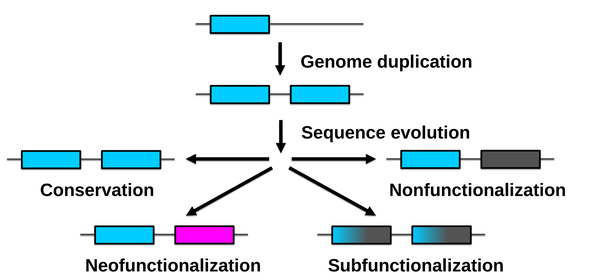

Figure 1: Functional evolutionary fates of duplicate genes.

ANCHORED HYBRID ELEMENTS:

- Gene regions under purifying selection
- ≥ 80% identity over ≥ 100 nucleotide regions
- Tailored to fit appropriate timescale
- Data was phased for four alleles
- Generated by Center for Anchored Phylogenomics (Lemmon *et al*., 2012)
- Outgroup sequences and probe design obtained from Stout *et al*. (2016)

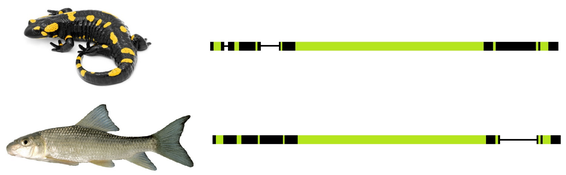

Figure 2: A depiction of anchored hybrid elements (AHEs) between a fish and a salamander. Green regions denote sites that are identical between the species, whereas black denotes regions that are variable. Anchoring sites are identified using a probe, followed by an outward sequencing into the more variable “flanking sites”.

METHODS:

- **Tree inference using Maximum Likelihood (GARLI)**

Maximum likelihood trees were generated for all 179 loci on the CIPRES phylogenetics portal using GARLI 2.01. All of the trees were assessed to determine which species were monophyletic for all four alleles. These data were recorded and analyzed by species, locus, and clade. These trees were also assessed to determine which loci had sequences from Ictiobinae falling into divergent clades.

- **Divergence dating using fossil calibrations (BEAST2)**

The 72 loci that had Ictiobinae sequences divided into two clades were analyzed using BEAST2 on CIPRES. These analyses were calibrated using three fossil calibrations; one for all Catostomidae (61.7 Mya) and one fossil for each of the two Ictiobinae clades (33.9 Mya). The BEAST analysis produced a Bayesian Inference tree containing node ages and 95% confidence intervals for each of the 72 loci.

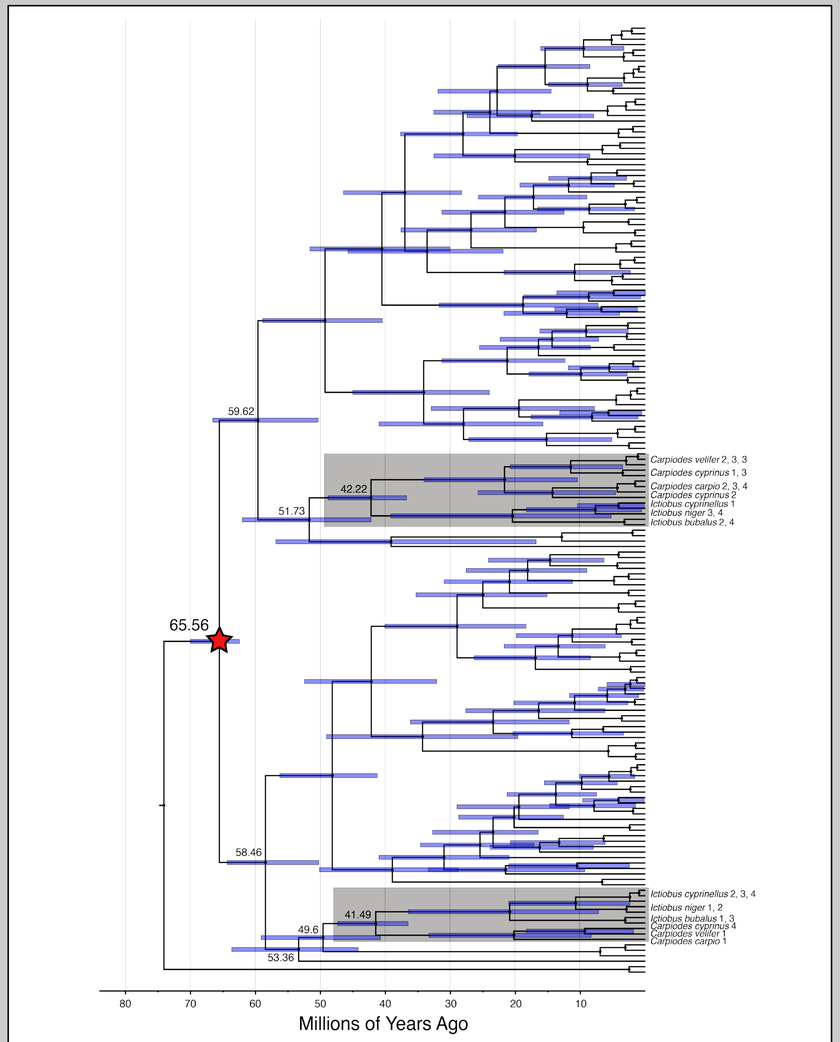

Figure 3: Example time-calibrated phylogeny. Highlighted in gray are two clades of Ictiobinae. Blue bars show the 95% highest posterior density (HPD) intervals of each node. Numbers are predicted node ages, and node marked with red star is the most recent common ancestor of the divergent clades.

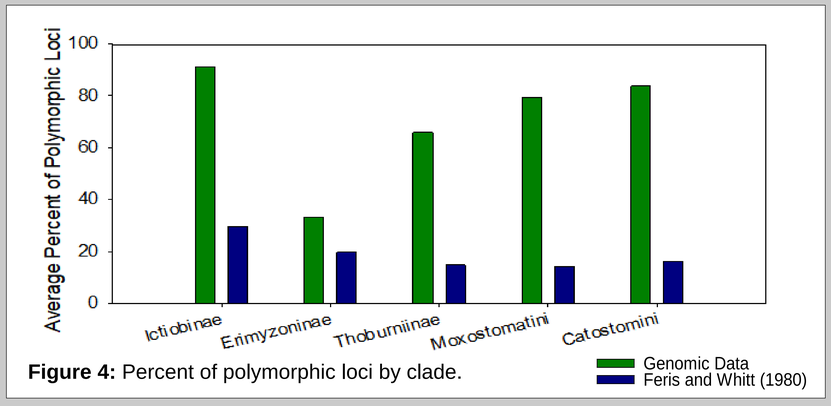

Figure 4: Percent of polymorphic loci by clade.

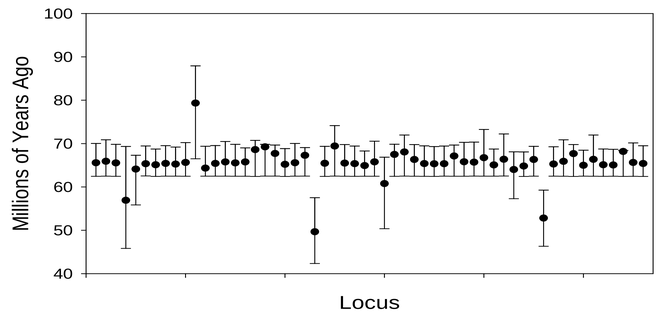

Figure 5: Age of node representing the most recent common ancestor of diverging clades of Ictiobinae. Circles show node age, and lines show the 95% HPD interval.

PRELIMINARY CONCLUSIONS:

- On average, species were divergent on 80% of the loci. Further analysis is necessary to determine the evolutionary fate of the duplicates (Fig. 5).
- Assessment of the maximum likelihood trees revealed that Ictiobinae has the highest percent of polymorphic loci out of all the subfamilies. This agrees with the results of Ferris and Whitt (1980), but trends of other subfamilies differ.
- Diverging alleles of Ictiobinae appear to have a most recent common ancestor between 60 and 70 Mya. This date indicates that the alleles diverged shortly after the WGD event that distinguishes the Catostomidae family.

LITERATURE CITED:

1. Ferris, Stephen D., and Gregory S. Whitt. "Genetic variability in species with extensive gene duplication: the tetraploid catostomid fishes." The American Naturalist 115.5 (1980): 650-666.
2. Lemmon, A. R., Emme, S. A., and Lemmon, E. M. 2012. "Anchored Hybrid Enrichment for Massively High-Throughput Phylogenomics." *Systematic Biology* 61.5: 727-44. Web.
3. Stout, C. C., Tan, M., Lemmon, A. R., Lemmon, E. M., Armbruster, J. W. 2016. “Resolving Cypriniformes relationships using an anchored enrichment approach”. *BMC Evolutionary Biology* 16:244. Web.

ACKNOWLEDGEMENTS:

CIPRES

IOWA EPSCoR

Dr. Gary and Myrna Floyd Undergraduate Research Fellowship